Computing

INTEGRATING BUSINESS & INFO...   WWW.BEYONDCOMPUTINGMAG.COM

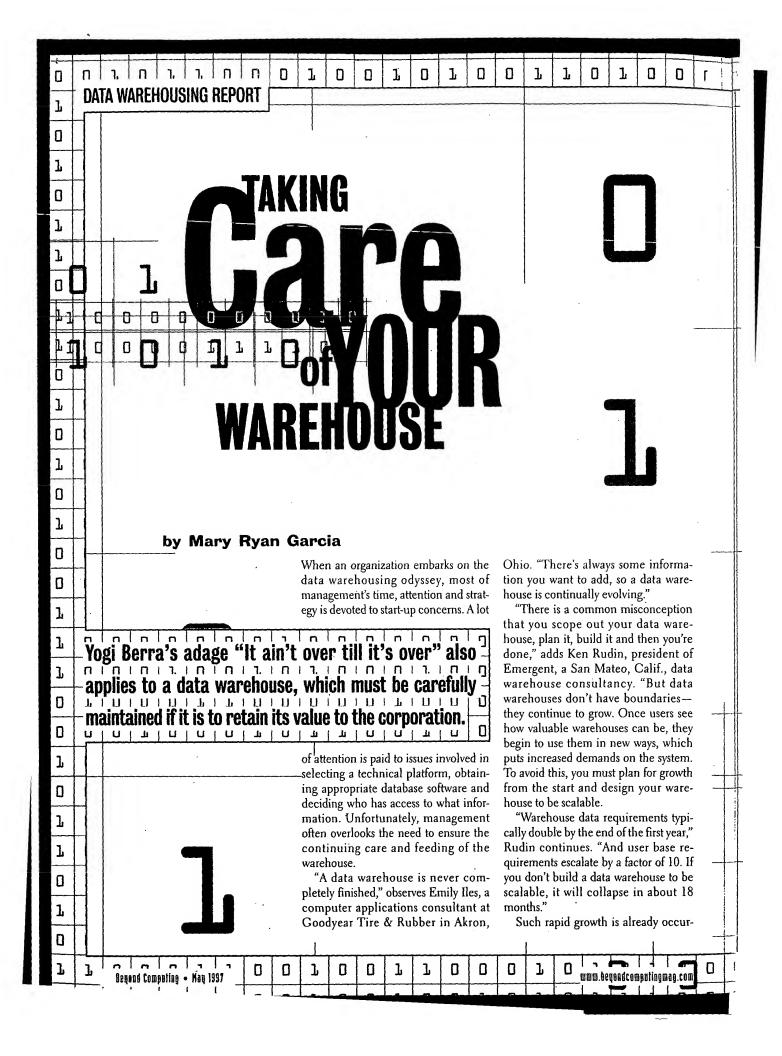## 21st Century Leadership
Do You Have What It Takes? p.38

DATA HOUSING
The Bottom Line
Decision Support
Maintaining Value

Legacy Meets the Net p.44,   Intranets p.50,   Teamwork p.52

TAKING Care of YOUR WAREHOUSE

by Mary Ryan Garcia

When an organization embarks on the data warehousing odyssey, most of management's time, attention and strategy is devoted to start-up concerns. A lot

Yogi Berra's adage "It ain't over till it's over" also applies to a data warehouse, which must be carefully maintained if it is to retain its value to the corporation.

of attention is paid to issues involved in selecting a technical platform, obtaining appropriate database software and deciding who has access to what information. Unfortunately, management often overlooks the need to ensure the continuing care and feeding of the warehouse.

"A data warehouse is never completely finished," observes Emily Iles, a computer applications consultant at Goodyear Tire & Rubber in Akron,

Ohio. "There's always some information you want to add, so a data warehouse is continually evolving."

"There is a common misconception that you scope out your data warehouse, plan it, build it and then you're done," adds Ken Rudin, president of Emergent, a San Mateo, Calif., data warehouse consultancy. "But data warehouses don't have boundaries—they continue to grow. Once users see how valuable warehouses can be, they begin to use them in new ways, which puts increased demands on the system. To avoid this, you must plan for growth from the start and design your warehouse to be scalable.

"Warehouse data requirements typically double by the end of the first year," Rudin continues. "And user base requirements escalate by a factor of 10. If you don't build a data warehouse to be scalable, it will collapse in about 18 months."

Such rapid growth is already occur-

KATHERINE LAMBERT FOR *BEYOND COMPUTING*

PHH plans to give clients who need
information on vehicle maintenance
access to its data warehouse via the
Internet, says

ring at PHH, a $5 billion provider of vehicle leases and residential services based in Hunt Valley, Md. The firm's data warehouse—a UNIX-based Sun Ultra Enterprise 5000 server from Sun Microsystems running Sybase SQL Server 11—debuted a year ago with 10 users and 40Gbytes of data. Since then, it has scaled to 200 users and 240Gbytes of data.

One key to maintaining this warehouse is its scalable design, says Mickey Lutz, a PHH director for technical planning and information management. "Unforeseen things always happen," he

histories. All data marts are integrated into the enterprise data warehouse, which is accessible by end users.

FEEDING TIME

In order to grow their data warehouses, companies need to "feed" them with data. To port data to the warehouse from its legacy systems—an Amdahl 5995-4550 mainframe and UNIX-based servers from Sun—PHH selected the ETI• EXTRACT Tool Suite from Evolutionary Technologies International. This enables PHH to automatically generate programs to selectively retrieve data from disparate

years. To ensure the accuracy of new data, the company is preserving the end user business team that judged the accuracy of the initial data when the system was developed during a six-week period last year.

The warehouse—developed using SAS/Warehouse Administrator running on an IBM PC Server 704—relied on these end users (called "data stewards") to compare the information in the SAS warehouse with data culled from IMS databases and SAS files on Goodyear's IBM mainframes, as well as with data from Lotus 1-2-3 spreadsheets housed on users' PCs.

"As the data stewards verify the data," Iles explains, "they begin to trust the numbers generated by the warehouse." She notes that it's crucial to employ end users to verify data because "these people are intimate with the data."

"I'm not a business expert," she continues. "I'm a person who takes business rules and applies them to the data warehouse to meet the data users' needs. I don't know what data the users need until they tell me, so it's important for them to be the owners of the system."

> "Organizations initially spend a lot of time ensuring the quality of the data....However, over time, the focus on the quality of data begins to fade."

says, "so the usage and design of the warehouse may have to change."

For example, new reporting requirements may come from internal and external sources. "Priorities of the business change over time," Lutz says. "We need to prepare for the unexpected."

The PHH warehouse has been phased in with a series of releases that began last May with major financial data. Additional releases have been adding further layers of data, such as detailed transaction records relating to PHH's fuel and vehicle maintenance cards. The latest release includes information on the vehicle maintenance assistance program.

Another firm that is rapidly deploying data warehousing is Irving, Texas-based GTE Supply, a division of telecommunications giant GTE that is responsible for supplying customers with communications equipment. Its warehouse—an IBM RS/6000 Scalable POWERparallel System running an Oracle 7.3 database—is now being accessed by 200 active users.

"Our plan is to roll out the data warehouse in 90-day increments," says Roger Copeland, administrator of new technology. "Each of our six data marts has been delivered incrementally." The data marts are centered on subject areas such as purchasing, inventory, measurements, planning systems and customer

database or file systems, and to validate and clean data to ensure accuracy.

"We generated 500,000 lines of error-free code in less than two months with only three programmers," Lutz reports. "We achieved at least a 10-to-1 productivity improvement." When you consider the fact that some programmers earn $100,000 or more a year, "it doesn't take long for the product to pay for itself," Lutz points out. "In our case, it took less than one year."

Another ongoing process involves maintaining data quality within a warehouse. According to a report by the Data Warehousing Institute, an industry research group in Gaithersburg, Md., data cleansing and network integration frequently take much longer than anticipated.

As Emergent's Rudin points out, "Organizations initially spend a lot of time ensuring the quality of the data, particularly data that overlaps subject areas, such as customer IDs or product IDs. However, over time, the focus on IDs. However, over time, the focus on the quality of data begins to fade."

As a result, data that is added later often isn't properly cleansed, Rudin notes. Eventually, the warehouse begins to accumulate "dirty data."

Goodyear doesn't plan to let that happen to its data warehouse as the number pen of people using it grows from two dozen to several hundred during the next few

GTE Supply also tries to prevent the problem of dirty data by deploying teams of end users who compare data in the warehouse with the information they get from the mainframe. It's a difficult task, Copeland points out, because each data center has between 10 and 30 separate systems.

Each "business-driver team"—which is composed of stakeholders and experts in a specific subject area—has the responsibility of validating the model, structure and content of its respective data mart.

The maintenance of a data warehouse extends down to the data that tracks the data: It's called metadata. "Metadata defines what's in your warehouse from different perspectives," explains Tricia Spencer, a principal with the Center for Advanced Technologies, the research arm of American Management Systems (AMS), a Fairfax, Va., IT consultancy.

From a technical perspective, database designers and administrators use metadata to capture the data type, source-system mapping, and transformation and cleansing rules. From a

business perspective, metadata provides end users with definitions, aliases, derivations, and predefined queries and reports.

"When you have a definition of what your warehouse contains, maintenance is easier because you have a record of what's there," Spencer says. "Successful organizations aren't static: They change data structures over time. With metadata, you have a history of the structure across time periods."

CROSSING PLATFORMS

Establishing and maintaining an enterprisewide data warehouse is time-consuming and may require a significant investment in a wide variety of tools for scalability, data mining and metadata on an ongoing basis.

One key to the successful implementation and maintenance of a data warehouse is to standardize software tools. "Data warehousing provides both a technique and a methodology to provide access to information across multiple platforms," notes GTE Supply's Copeland.

Throughout its data warehouse, GTE utilizes the FOCUS Six Managed Reporter Edition from Information Builders. It offers a client/server architecture that distributes application logic, business logic and presentation among multiple computer platforms.

"Every day, we extract a subset of information from the mainframe, scrub it and bring it into Oracle," Copeland reports. The firm uses Information Builder's Enterprise Copy Manager, a data extraction and scheduling tool that maintains daily records, keeps a rolling 36 months of history, adds data and purges old data.

Because warehouses often must support multiple platforms or remote access, the Internet is beginning to play an important role in allowing users access to a data warehouse. According to Lutz, PHH will be giving its clients Internet access to vehicle maintenance

data in its warehouse this summer. And Copeland is enthusiastic about Internet connectivity for GTE Supply's warehouse.

"We have a number of end users at remote sites," he reports. "Currently, end users can generate *ad hoc* reports through the Synergy website, which is an intranet website that is available to all GTE Supply employees."

GTE Supply uses Information Builder's WebFOCUS in its warehouse to allow end users to create true *ad hoc* queries directly from the data warehouse using a Microsoft NT Server version 4.0. The company hopes to further enhance the warehouse with Java applications, including web-based cataloging and other electronic capabilities such as graphing and online order tracking. Copeland notes that appropriate maintenance of an enterprisewide data warehouse depends on the selection of middleware.

"Middleware is the nervous system of the data warehousing architecture," Copeland explains. "It is imperative that this system be easy to configure and administer in order to provide transparent usage of the warehouse by users."

Data warehouses will continue to expand in order to accommodate the information requirements of today's organizations. And that means that companies must devote serious attention to their continued care and feeding to ensure healthy transitions throughout the various stages of growth.

Mary Ryan Garcia *is a freelance technology journalist based in Coram, N.Y.*

> ## Establishing and maintaining an enterprisewide data warehouse is time-consuming and may require a significant investment in a wide variety of tools....

**PHH Vehicle
Management Services**

PHH

1 800 ONLY PHH
www.phh.com

United States • Canada • United Kingdom • Ireland • Germany • France • Belgium • Italy • Netherlands • Spain